

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 10/528,872  
Source: PCT/10  
Date Processed by STIC: 4/6/05

# ***ENTERED***



PCT

## RAW SEQUENCE LISTING

DATE: 04/05/2005

PATENT APPLICATION: US/10/528,872

TIME: 12:11:19

Input Set : A:\PTO.AMC.txt

Output Set: N:\CRF4\04052005\J528872.raw

3 <110> APPLICANT: DSM IP ASSETS B.V.  
 5 <120> TITLE OF INVENTION: SQS gene  
 7 <130> FILE REFERENCE: NDR5218  
 C--> 9 <140> CURRENT APPLICATION NUMBER: US/10/528,872  
 C--> 9 <141> CURRENT FILING DATE: 2005-03-23  
 9 <150> PRIOR APPLICATION NUMBER: PCT/EP03/10573  
 10 <151> PRIOR FILING DATE: 2003-09-23  
 12 <150> PRIOR APPLICATION NUMBER: EP 02021619.8  
 13 <151> PRIOR FILING DATE: 2002-09-27  
 15 <160> NUMBER OF SEQ ID NOS: 8  
 17 <170> SOFTWARE: PatentIn version 3.2  
 19 <210> SEQ ID NO: 1  
 20 <211> LENGTH: 4807  
 21 <212> TYPE: DNA  
 22 <213> ORGANISM: Phaffia rhodozyma  
 25 <220> FEATURE:  
 26 <221> NAME/KEY: 5'UTR  
 27 <222> LOCATION: (1469)..(1470)  
 29 <220> FEATURE:  
 30 <221> NAME/KEY: exon  
 31 <222> LOCATION: (1550)..(1577)  
 33 <220> FEATURE:  
 34 <221> NAME/KEY: Intron  
 35 <222> LOCATION: (1578)..(1752)  
 37 <220> FEATURE:  
 38 <221> NAME/KEY: exon  
 39 <222> LOCATION: (1753)..(1766)  
 41 <220> FEATURE:  
 42 <221> NAME/KEY: Intron  
 43 <222> LOCATION: (1767)..(1882)  
 45 <220> FEATURE:  
 46 <221> NAME/KEY: exon  
 47 <222> LOCATION: (1883)..(2071)  
 49 <220> FEATURE:  
 50 <221> NAME/KEY: Intron  
 51 <222> LOCATION: (2072)..(2182)  
 53 <220> FEATURE:  
 54 <221> NAME/KEY: exon  
 55 <222> LOCATION: (2183)..(2397)  
 57 <220> FEATURE:  
 58 <221> NAME/KEY: Intron  
 59 <222> LOCATION: (2398)..(2474)  
 61 <220> FEATURE:

p.6

## RAW SEQUENCE LISTING

DATE: 04/05/2005

PATENT APPLICATION: US/10/528,872

TIME: 12:11:19

Input Set : A:\PTO.AMC.txt

Output Set: N:\CRF4\04052005\J528872.raw

```

62 <221> NAME/KEY: exon
63 <222> LOCATION: (2475)..(3087)
65 <220> FEATURE:
66 <221> NAME/KEY: Intron
67 <222> LOCATION: (3088)..(3230)
69 <220> FEATURE:
70 <221> NAME/KEY: exon
71 <222> LOCATION: (3231)..(3356)
73 <220> FEATURE:
74 <221> NAME/KEY: Intron
75 <222> LOCATION: (3357)..(3453)
77 <220> FEATURE:
78 <221> NAME/KEY: exon
79 <222> LOCATION: (3454)..(3475)
81 <220> FEATURE:
82 <221> NAME/KEY: Intron
83 <222> LOCATION: (3476)..(3564)
85 <220> FEATURE:
86 <221> NAME/KEY: exon
87 <222> LOCATION: (3565)..(3881)
89 <220> FEATURE:
90 <221> NAME/KEY: Intron
91 <222> LOCATION: (3882)..(3958)
93 <220> FEATURE:
94 <221> NAME/KEY: exon
95 <222> LOCATION: (3959)..(3970)
97 <220> FEATURE:
98 <221> NAME/KEY: polyA_site
99 <222> LOCATION: (4106)..(4107)
101 <400> SEQUENCE: 1
102 gttcctgttc agtcaaagag tgggaaaaaac atgaaagtaa aaagatgtaa tgaaagaagg      60
104 ggtcagaaca tcggagatac aatggcccat agaggaagga aagctactta ccagaaacca      120
106 gtgaggtttg cctaggaagt aatcccttcg tttctcaaag atatcttttt tgaaagcatc      180
108 gatgaacgac atgtcgaaac catctccatc ctcgaaatca agtttactcg atttagacct      240
110 ttccagcttt tctgctctct ccagtttcgc agctttctct tcgggaagaa gctctccgcc      300
112 agtcgatgtt ctgtcgacag gagaccagta gaaggcggaa ccgacaattt tggatggatc      360
114 ggaggacagg gtggctttta caaatcggtg gtacggagga tcgaacggcg cttctctcgt      420
116 tcgaaggttg actcctcttg ctatgtgtat gagagcatat ccgttgatgt ctcagttaaa      480
118 atttcttttt ctttctaccc ggagagtaag acacacaaag aatcacgaag aatatgatga      540
120 ctgaccgatc cgaatatcta gcgcagggtt cttctctact ggttccattc ttcgaacgat      600
122 aggttcatgt ttgaaagcat tgatcctagt tgccctctatc tgaggccagt ctgccaatgt      660
124 agcaggctca atgatcactt ggggtttgtg catcttgatg ttcaaccaag tgcgcaacg      720
126 gtcgagattc ttttttcttc ttttggtcga gaaaaaaaaa cggcttcgct tcgcacgcgc      780
128 gcggggggatc acccgcatat taagcgggat gacgctcatc aaccggccaa gtgttcttca      840
130 tcataggtga aggttaaaac ggaatggata ggaggagcta accacgtttt tattttaatt      900
132 cgacttgggc agcctcgtec atagtgtctg atggttatat cgcatagaa aggcagcgcc      960
134 tggcgggttc gtcatggccg tgatcatctg ctttgttaga cattgtccat cagtccctc     1020
136 aatgacagtt tcccagcgcc atcactaaga cacaaacgta tccagcacgc catgtccatc     1080
138 actgaagaag gtagggctct gtcgagccag tgcaaccaga gttacagatg aacatcaggc     1140

```

## RAW SEQUENCE LISTING

DATE: 04/05/2005

PATENT APPLICATION: US/10/528,872

TIME: 12:11:19

Input Set : A:\PTO.AMC.txt

Output Set: N:\CRF4\04052005\J528872.raw

```

140 cttgatacaga cccgacttat gaatatggcc gttattgtac acttcttggt gtcctcagag 1200
142 ctgctcttttc gtgttttttca ctttctttcc ggatcaaacg agactgctcg tgtatctatc 1260
144 tgtgcttgcc atatgagcat cccatgcctc tgctcaaagt atgctggagc tacgatccat 1320
146 cagagacgac acaaaacggg gttgtatgaa ctctacattt cctaagtta ttggaatttt 1380
148 ctgtaatgag ttcttcatct ttctctaatg cttttttgta gtccgtcttt tcaaccttgc 1440
150 cagcgttttc cgtgtcttct ttctcctttg acggtcacat ctttcttctc tcttctcggt 1500
152 ctttcttccg tccttctctc tctctcttcg tctgaacatc agcatcctc atg ggc ata 1558
153 Met Gly Ile
154 1
156 tca gat tac ctc gtt ctg g gtcagttctg tcttttggtt gattcttctc 1607
157 Ser Asp Tyr Leu Val Leu
158 5
160 ttcttgccgg cggtcgctg tcttgggtat atcatcagca atgagaaaca tgatgttccc 1667
162 cccgcgtcaa tcaactgacct ttgtgctctc tacttctttc ctgtcgaatt gatcctgatt 1727
164 gatacgtgtg ccggtgctt aacag ct ttc acg cat cct gtaggtgttt 1776
165 Ala Phe Thr His Pro
166 10
168 tatcgtatgc ttcattgtga tgtttagtca cgcggactga cctggccggg tgattttctg 1836
170 tatgatcgct tgtgctaccg tctttcttgg aaatccttcc catcag gcc gat ctg 1891
171 Ala Asp Leu
172 15
174 cga gct tta atg cag tac gcg atc tgg cat gag cct cga agg aat atc 1939
175 Arg Ala Leu Met Gln Tyr Ala Ile Trp His Glu Pro Arg Arg Asn Ile
176 20 25 30
178 act gca cag gag gaa cat gca aca tcc ggt tgg gac cga gaa act atg 1987
179 Thr Ala Gln Glu Glu His Ala Thr Ser Gly Trp Asp Arg Glu Thr Met
180 35 40 45
182 aag gaa tgt tgg aag tat ttg gat ctg act tca aga agt ttc gca gct 2035
183 Lys Glu Cys Trp Lys Tyr Leu Asp Leu Thr Ser Arg Ser Phe Ala Ala
184 50 55 60 65
186 gtc atc aaa gag ttg gac gga gat ctt acc cga gtc gtacgtgttt 2081
187 Val Ile Lys Glu Leu Asp Gly Asp Leu Thr Arg Val
188 70 75
190 tcattcttctc tctcctttga gatctggctg cctccgcatt ttcttggtgc agaaggggtca 2141
192 gaagctgaca acaccatctc tactgttcgg gacacggcta g atc tgt tta ttc tat 2197
193 Ile Cys Leu Phe Tyr
194 80
196 ctc gct ctt cga gga ctg gat acc att gag gat gac atg agt cta tct 2245
197 Leu Ala Leu Arg Gly Leu Asp Thr Ile Glu Asp Asp Met Ser Leu Ser
198 85 90 95
200 aat gat gtg aag ctt ccc ctg ctt cgg aca ttc tgg gaa aag ctt gac 2293
201 Asn Asp Val Lys Leu Pro Leu Leu Arg Thr Phe Trp Glu Lys Leu Asp
202 100 105 110
204 tcc cct ggg tgg acc ttt act gga tcc ggt cca aat gag aag gat aga 2341
205 Ser Pro Gly Trp Thr Phe Thr Gly Ser Gly Pro Asn Glu Lys Asp Arg
206 115 120 125 130
208 gag ctt ctt gtt cac ttc gat gtg gcc atc gcc gag ttt gcc aac ttg 2389
209 Glu Leu Leu Val His Phe Asp Val Ala Ile Ala Glu Phe Ala Asn Leu
210 135 140 145

```

## RAW SEQUENCE LISTING

DATE: 04/05/2005

PATENT APPLICATION: US/10/528,872

TIME: 12:11:19

Input Set : A:\PTO.AMC.txt

Output Set: N:\CRF4\04052005\J528872.raw

```

212 gac gtc aa gtgagtttcc ctttatgggtt ggatcatccg ctcgacagac 2437
213 Asp Val Asn
216 tcgaaacgct catcactttg gtctgcttga tgaacag c tct cgg aac gtc att 2490
217 Ser Arg Asn Val Ile
218 150
220 cga gac atc act cgc aag atg ggt aac ggt atg gcc gac ttt gct tct 2538
221 Arg Asp Ile Thr Arg Lys Met Gly Asn Gly Met Ala Asp Phe Ala Ser
222 155 160 165 170
224 ctc tct acg ccc tcc aag cct gtg gcc gag gtc cag tcg acc gaa gat 2586
225 Leu Ser Thr Pro Ser Lys Pro Val Ala Glu Val Gln Ser Thr Glu Asp
226 175 180 185
228 ttc aac cta tac tgt cat tac gtc gct gga ctc gtc ggc gag gga ctc 2634
229 Phe Asn Leu Tyr Cys His Tyr Val Ala Gly Leu Val Gly Glu Gly Leu
230 190 195 200
232 tcc cga ctc ttt gtc gcg acc gag aag gaa cga cca ttc ttg gcc aac 2682
233 Ser Arg Leu Phe Val Ala Thr Glu Lys Glu Arg Pro Phe Leu Ala Asn
234 205 210 215
236 cag atg gta ctt tca aac tcg ttc gga ctc ctt ctc caa aag aca aac 2730
237 Gln Met Val Leu Ser Asn Ser Phe Gly Leu Leu Leu Gln Lys Thr Asn
238 220 225 230
240 atc ctt cga gat att cgg gag gac gcc gac gaa ggt cgt ggc ttc tgg 2778
241 Ile Leu Arg Asp Ile Arg Glu Asp Ala Asp Glu Gly Arg Gly Phe Trp
242 235 240 245 250
244 cca aga gag atc tgg gcc aac ccg atc tat act gcg cat gca ccg ggc 2826
245 Pro Arg Glu Ile Trp Ala Asn Pro Ile Tyr Thr Ala His Ala Pro Gly
246 255 260 265
248 aca agg ttt aac tcg ttg act gac ctg gtc aag aaa gaa aac atc gac 2874
249 Thr Arg Phe Asn Ser Leu Thr Asp Leu Val Lys Lys Glu Asn Ile Asp
250 270 275 280
252 aaa gga tca atg tgg gtg ttg agt gcg atg aca ctc gac gcg atc acc 2922
253 Lys Gly Ser Met Trp Val Leu Ser Ala Met Thr Leu Asp Ala Ile Thr
254 285 290 295
256 cat act acc gac gca ctg gac tac ctc tca ctt cta aag aac cag agt 2970
257 His Thr Thr Asp Ala Leu Asp Tyr Leu Ser Leu Leu Lys Asn Gln Ser
258 300 305 310
260 gtt ttc aac ttt tgt gct atc ccg gct gtc atg tcg att gca acg ttg 3018
261 Val Phe Asn Phe Cys Ala Ile Pro Ala Val Met Ser Ile Ala Thr Leu
262 315 320 325 330
264 gag cta tgc ttc atg aac cca gcg gtg ttc caa cga aac ata aaa atc 3066
265 Glu Leu Cys Phe Met Asn Pro Ala Val Phe Gln Arg Asn Ile Lys Ile
266 335 340 345
268 aga aag gga gaa gcc gtc gag gtgcgttcgc gcgttctgtt tctacctttc 3117
269 Arg Lys Gly Glu Ala Val Glu
270 350
272 ataacattgg aggttcttga ctcttaagcg tcttccaatc tgatgcctcc aattatcatc 3177
274 atttttgtct tttttgcttt cctcttggtt cttttcggcg tgattcaatc cag ctc 3233
275 Leu
278 att atg aag tgc aac aac cct cgg gag gtg gca tac atg ttt aga gat 3281
279 Ile Met Lys Cys Asn Asn Pro Arg Glu Val Ala Tyr Met Phe Arg Asp

```

## RAW SEQUENCE LISTING

DATE: 04/05/2005

PATENT APPLICATION: US/10/528,872

TIME: 12:11:19

Input Set : A:\PTO.AMC.txt

Output Set: N:\CRF4\04052005\J528872.raw

```

280 355          360          365          370
282 tat gct cga aag att cat gcc aag gct att cct aca gat cct aac ttc      3329
283 Tyr Ala Arg Lys Ile His Ala Lys Ala Ile Pro Thr Asp Pro Asn Phe
284          375          380          385
286 atc aag ttg agc gtt gcg tgt ggt cga gtgagttgat cgatecatcc      3376
287 Ile Lys Leu Ser Val Ala Cys Gly Arg
288          390          395
290 atctttttgtt ttgatcatcg cgagacttga ctgacgcatt actcaaaaca tcatcgcttc      3436
292 tccttcttgc tctctag atc gaa caa tgg gct gag cac t gtatgttcct      3485
293          Ile Glu Gln Trp Ala Glu His
294          400
296 ccgccccctcc ttcaagtttc ctctcgcttc atctttgttg agaagaggga tctgatgtat      3545
298 ctttctttgt tcggatcag ac tac ccc tca ttt atg atg att cgg cct tcg      3596
299          Tyr Tyr Pro Ser Phe Met Met Ile Arg Pro Ser
300          405          410
302 aat gac cct caa aac ccc gca ccc tca acg gcg ctt gac cct ttc tca      3644
303 Asn Asp Pro Gln Asn Pro Ala Pro Ser Thr Ala Leu Asp Pro Phe Ser
304          415          420          425
306 gga gac gct cgt tta agg ata gcc tct aag aag gct gag atc acc gcc      3692
307 Gly Asp Ala Arg Leu Arg Ile Ala Ser Lys Lys Ala Glu Ile Thr Ala
308 430          435          440          445
310 gct gct ctt gtc agg aag aaa gcc cgg gat cac gct aag tgg aga gag      3740
311 Ala Ala Leu Val Arg Lys Lys Ala Arg Asp His Ala Lys Trp Arg Glu
312          450          455          460
314 tcc aag gga ttg cct ccg agc gat ccg aac aag ccg gac aac tcg gag      3788
315 Ser Lys Gly Leu Pro Pro Ser Asp Pro Asn Lys Pro Asp Asn Ser Glu
316          465          470          475
318 gat gtt aat tgg gta ttg atc ggc ggt atg atc gtt gga ttg ttg ctc      3836
319 Asp Val Asn Trp Val Leu Ile Gly Gly Met Ile Val Gly Leu Leu Leu
320          480          485          490
322 gtg atg ggc gtg ctc ggt ttg gct atc gct tgg gtt gtt ctt cag      3881
323 Val Met Gly Val Leu Gly Leu Ala Ile Ala Trp Val Val Leu Gln
324          495          500          505
326 gtgcgttctt ccaaagagcc tttctctcat gaacacgcac ataggttgat ctaattctat      3941
328 cttactctgt catacag ttt gag caa taa tctcaagatt ctagtccatc      3990
329          Phe Glu Gln
330          510
332 ctttcgctca acgatctgct tcttctcctt ctcttctctc gtcttctctg gtttcttttc      4050
334 ttactttctg ggatcttcct tcttgaatcc tccgatccaa tgtaatctgc ataccctcgc      4110
336 tttagtagaa accgatcctt cattcgatct tggcgaaaat ctaagcaaag agaatcactt      4170
338 ttgtctaata aaatttcctt taaagagtcg gctttttctt gtggcgaagc ttcaccccg      4230
340 ctctctctgg accatctctt ctcaatattc tttgtgctac tatatgatca agttctttga      4290
342 aatcaaagaa gaacatgtat ttgattttga ggttccaaga atacaaccgg cccaagtcgt      4350
344 tcttcgcagt tttcatcaga cagcacatat ctctcctcct ctctatagaa gccgatggg      4410
346 gccaatcgac tctcatgggt agaccgtgcc cttttgacac ggggagaaaag agaacgaaaag      4470
348 gacacttgac cgattcggtt ataaagccgt cccacacctt tctttaatgg caattcaaga      4530
350 agagaaaaac aaccctcgcg cgactcgag tagtcgatca gaccttccga acgacagata      4590
352 tcatttgctg aaatcgaccg gattttaaag ctgctgccag gtcggtgaat cccctaggt      4650
354 gatctccttg tacaaagatg ttgggcacgg acttttcgac ccggatgaga acgtcgtgaa      4710

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/10/528,872

DATE: 04/05/2005  
TIME: 12:11:20

Input Set : A:\PTO.AMC.txt  
Output Set: N:\CRF4\04052005\J528872.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:4; N Pos. 3,6,12,15  
Seq#:5; N Pos. 3,9,15,18,21  
Seq#:6; N Pos. 3,6,9,12,15

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete,  
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:4,5,6,7,8

**VERIFICATION SUMMARY**

PATENT APPLICATION: **US/10/528,872**

DATE: 04/05/2005

TIME: 12:11:20

Input Set : **A:\PTO.AMC.txt**

Output Set: **N:\CRF4\04052005\J528872.raw**

L:9 M:270 C: Current Application Number differs, Replaced Current Application No

L:9 M:271 C: Current Filing Date differs, Replaced Current Filing Date

L:671 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:0

L:709 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5 after pos.:0

L:747 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6 after pos.:0